心理学报 2018, Vol. 50, No.4, 400−412
Acta Psychologica Sinica DOI: 10.3724/SP.J.1041.2018.00400

频率树类型和提问方式对因果强度估计模式的影响*

刘雁伶[1] 陈 军[2] 沈友田[3] 胡竹菁[3]

([1]江西科技师范大学教育学院，南昌 330038) ([2]中国科学院心理研究所，北京 100101)
([3]江西师范大学心理学院，江西省心理与认知科学重点实验室，南昌 330022)

摘 要 心理学研究的重要目的之一发现心理干预的途径和方法。但截至目前，有效干预人类被试因果推理过程的手段尚不丰富，干预手段的效果并不稳定。本研究采用完全随机设计开展两个实验，分别探讨频率树是否影响大学生被试在反事实提问和能力提问因果推理问题上的作答表现。结果显示：(1)在两个实验中都发现了明显的图形促进效应，大部分被试在借助提供嵌套集合关系频率树(而非隐藏嵌套集合关系频率树)辅助推理时使用 PPC 值估计因果强度；(2)频率树类型和提问方式共同影响被试的因果强度估计模式，提供嵌套集合关系频率树+反事实提问的组合促使最多被试使用 PPC 估计因果强度。结果说明：明确数据之间的嵌套集合关系能极大地提高被试使用 PPC 估计因果强度的概率，关注焦点集信息有助于被试明确数据间的嵌套集合关系。

关键词 因果推理；图形促进效应；频率树；提问方式
分类号 B842

1 问题提出

最近 30 年，因果推理研究逐渐成为认知科学研究的核心领域之一(Holyoak & Cheng, 2011)，研究焦点集中于确定导致效果的原因(Lombrozo, 2012; Powell, Merrick, Lu, & Holyoak, 2016)、判断因果结构特点(Griffiths & Tenenbaum, 2005; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008; Cheng, Liljeholm, & Sandhofer, 2013)和估计因果之间的联结强度(Cheng, 1997; 王墨耘，傅小兰，2004a; Luhmann & Ahn, 2011; Yeung & Gtriffiths, 2015; Buchsbaum, Gtriffiths, Plunkett, Gopnik, & Baldwin, 2015)，形成了多种解释被试因果推理表现的模型(Jenkins & Ward, 1965; Rescorla & Wagner, 1972; Cheng, 1997; Lu et al., 2008; Carroll, Cheng, & Lu, 2013)，其中影响较大的有 Jenkins 和 Ward (1965)提出的概率对比模型(Probabilistic contrast model, 以下简记为 ΔP 模型)和 Cheng (1997)在修正概率对比模型基础上提出的因果力理论(the causal Power theory of Probabilistic Contrast model, 以下简记为 PPC 模型)。ΔP 模型通过比较因果协变关系(见表 1)中目标原因出现前后效果出现的概率差来描述因果之间的联结强度，采用表 1 的符号，$\Delta P = a/(a + b) - c/(c + d)$。ΔP 为正表示目标原因能导致效果(协变关系被称为产生式)，ΔP 为负表示目标原因能消除效果(协变关系被称为预防式)，ΔP 的绝对值越大说明目标原因对效果的影响力越强；PPC 模型假定背景原因一直存在并导致效果，认为被试基于目标原因对焦点集的影响来判断因果结构和估计因果强度，产生式焦点集是表 1 中的 d，预防式焦点集是表 1 中的 c。PPC 模型分别使用 $\Delta P/(d/(c + d))$和$-\Delta P/(c/(c + d))$来预测被试对产生式和预防式协变关系的强度估计。

目前关于因果强度估计的研究大多着眼于描述和解释被试的因果推理表现，但心理学的研究不

收稿日期: 2016-11-09
* 国家自然科学基金(31460252)、江西省社会科学十三五规划课题(16JY17)、江西科技师范大学博士科研启动基金(3000990102)资助。
通信作者: 胡竹菁, E-mail: huzjing@ jxnu.edu.cn

表 1　目标原因和效果之间的协变关系

目标原因状态	效果出现	效果不出现
目标原因出现	a	b
目标原因不出现	c	d

注：协变关系部件 a、b、c、d 分别指代一次观测中目标原因出现效果也出现、目标原因出现效果不出现、目标原因不出现效果出现、目标原因不出现效果不出现的频数，在目标原因不出现时效果也可能出现(c 大于 0)，此时导致效果的原因被统称为背景原因。

仅要描述和解释心理现象，还要发现心理干预的途径和方法。从这个角度看，现有研究对如何干预被试，提高被试强度估计合理性的关注是不够的。提高因果强度估计的合理性需要解决两个问题：一是何谓合理的因果强度估计？二是如何帮助被试达成合理的因果强度估计？第一个问题的答案似乎已有共识：PPC 模型的预测被认为是因果强度的极大似然估计(Cheng, 1997; Griffiths & Tenenbaum, 2005; Holyoak & Cheng, 2011)，本研究也将被试使用 PPC 值估计因果强度视为合理的强度估计模式；对于第二个问题，有一些研究进行了初步探讨：虽然 PPC 模型的预测被认为是因果强度的极大似然估计，但是被试的强度估计往往偏离PPC模型的预测，Lober 和 Shank (2000)使用能力提问方法(以目标原因是服用某种药物，效果是病人头痛为例，能力提问方法指向被试提问：药物导致病人头痛症状的能力如何？)开展的研究发现当协变关系的 PPC 值不变时，被试的因果强度估计随 ΔP 值的变化而变化，Buehner, Cheng 和 Clliford 等(2003)认为提问方式是导致这些偏离的原因，他们使用反事实提问(反事实提问指向被试提出问题：现有 100 名病人在服用药物前不头痛，这些病人在服用该药物后有多少人会头痛？)的研究发现 80%的被试在多数情况下使用 PPC 作答。但 Perales 和 Shanks (2008)的实验 3b 却显示在反事实提问条件下没有一名被试真正使用 PPC 作答；反事实提问条件下的因果强度估计还受协变关系(Liljeholm & Cheng, 2009)、因果方向(Lu et al., 2008)和实质取样数量(即焦点集大小)的影响(Liljeholm & Cheng, 2009, 刘雁伶, 陈水平, 胡竹菁, 2015)的影响，部分被试的作答并不遵循 ΔP 或 PPC 的预测，而是落在 ΔP 和 PPC 的预测值之间(Cheng & Buehner, 2012)。总之，现有研究显示反事实提问不能保证大多数被试使用 PPC 值估计因果强度，是否存在促使被试使用PPC值估计因果强度的其他方式？

对人的推理过程进行适当干预可以有效提高人们推理的正确率，有助于人们在工作和生活中避免推理偏差。一个成功的例子来自贝叶斯推理领域：该领域的研究发现使用明确嵌套集合关系的频率树表征(Gigerenzer & Hoffrage, 1995; 史滋福, 邱江, 张庆林, 2006; Sirota, Kostovičová, & Vallée-Tourangeau, 2015)和解释(Brase, 2014)贝叶斯问题能显著提高被试正确解答贝叶斯问题的概率，研究成果被广泛应用于医疗、管理、法律和教育(Hill & Brase, 2012; Hoffrage, Hafenbrädl, & Bouquet, 2015)等领域。频率促进效应的原因可能是频率树将数据间的嵌套集合关系形象地提供给被试(Hoffrage et al., 2015; Vallée-Tourangeau, Abadie, & Vallée-Tourangeau, 2015)，基于同样的原因，频率表征还可以促进被试解决三段论推理(Thompson & Evans, 2012)和概率推理(Oaksford & Chater, 2010)问题。在因果强度估计任务中，被试使用PPC作答需要明确焦点集大小和目标原因在焦点集中引起效果的数量，其中焦点集是因果协变数据的子集，目标原因在焦点集中引起效果的部分是焦点集的子集，如果频率树能形象地表征数据间的嵌套集合关系，就有理由推测频率树能提高被试使用 PPC 值估计因果强度的概率。

Vallée-Tourangeau, Payton 和 Murphy (2008)使用文字、表格和频率树三种形式呈现协变关系，发现频率树能帮助被试区分不同强度的因果关系，但该研究直接移植了贝叶斯推理问题的频率树，并未明确因果协变数据间的嵌套集合关系，研究结果也不能说明频率树是否提高了被试使用 PPC 估计因果强度的概率。该领域的研究需澄清两个问题：一是明确嵌套集合关系的频率树是否能提高被试因果强度估计的合理性？对这一问题的探讨有利于发掘干预被试的有效方法；二是因果强度估计有几种典型的作答模式？分析被试的作答模式是推理心理学研究的重要内容，可以为避免错误推理模式提供建议。但因果强度推理领域的大部分研究使用所有被试的因果强度估计平均数为因变量，并不注重分析每名被试的作答依据和思维过程(相关批评也见王墨耘, 傅晓兰, 2004a)，也就无法描述和解释被试的作答模式。为回答以上两个问题，本研究设计适合因果推理任务的频率树，开展两个实验探究频率树对因果强度估计的促进效应，分析最典型的作答模式，同时考察提问方式、因果方向对被试因果推理作答模式的影响，为提高被试的因果强度估计合理性提供建议。

研究假设包括：研究显示明确嵌套集合关系是

图形促进效应的关键(史滋福等, 2006; Thompson & Evans, 2012; Sirota et al., 2015)。在因果推理中使用 PPC 估计因果强度需要被试明确协变关系总次数、焦点集次数和目标原因在焦点集中引起效果的数量三者之间的子集关系，如果频率树能形象地表征因果协变的子集关系，就有理由推测频率树能提高被试使用 PPC 估计因果强度的概率。所以有假设 1：在所有实验条件下，提供嵌套集合关系的频率树比隐藏嵌套集合关系的频率树促使更多被试使用 PPC 估计因果强度。

如前所述，现有研究大多以所有被试的因果强度估计平均数为因变量，较少对作答模式的研究，因此被试在不同协变关系上的作答模式是否存在差异的实验证据还较少。Buehner 等(2003)发现 80%的被试在多数情况下使用PPC估计因果强度，同时存在其他的估计模式，但该研究未报告被试在具体协变关系上的作答模式差异。现有仅见 Perales 和 Shanks (2008)发现被试在 ΔP 值很小(= 0.1)且 PPC 值中等(= 0.5)的协变关系上并不使用 PPC 估计因果强度，在 ΔP 值中等(= 0.45)或较大(= 0.7)的协变关系上被试基于 ΔP、PPC 和 P (E/C)估计因果强度。本研究拟比较被试在 ΔP 小(= 0.33) PPC 中等(= 0.5)、ΔP 小(= 0.33) PPC 大(= 0.83)、ΔP 大(= 0.67) PPC 大(= 0.83)协变关系上的作答模式差异，为该领域研究提供更多的实证数据。结合研究假设 1：提供嵌套集合关系的频率树可以增加被试在所有协变关系上使用PPC估计因果强度的概率，从而缩小不同协变关系上的估计模式差异，所以有假设 2：在使用提供嵌套集合关系的频率树辅助推理时，被试在不同协变关系上的强度估计模式不存在显著差异；在使用隐藏嵌套集合关系的频率树辅助推理时，被试在不同协变关系上的作答模式存在差异，具体的差异形式有待观察。

另外，被试的因果强度估计模式还受因果方向的影响。王墨耘和傅小兰(2004a, b)发现与产生式协变关系相比，更多被试在预防式协变关系上使用 PPC 估计因果强度；Lu 等(2008)预测一个预防式优势：被试在 ΔP 小(= 0.25)且 PPC 最大(= 1)的预防式协变关系上的强度估计要高于在相应产生式协变关系上的强度估计(根据定义，一个协变关系的 PPC 值一定不小于 ΔP 值，预防式优势暗示更多被试在预防式问题上使用 PPC 估计因果强度)，但他们的实验仅发现预防式优势在总体上是存在的，并未报告这一优势是否出现在单个协变关系上(相似的结果也见 Liljeholm & Cheng, 2009)；Yeung 和 Griffiths (2015)提出的实验先验模型则预测因果方向对被试强度估计的影响不如 Lu 等(2008)预测的那么大，被试在预防式协变关系的强度估计只是略高于产生式协变关系上的强度估计。综合以上研究结果，本研究认为较合理的假设是因果方向对强度估计的影响在总体上是存在，但就具体的协变关系而言，因果方向的变化并不导致被试的估计模式发生显著变化。结合研究假设 1，提供嵌套集合关系的频率树可以提高被试在所有协变关系上使用 PPC 估计因果强度的概率，从而缩小因果方向导致的作答模式差异，所以有研究假设 3：在使用提供嵌套集合关系的频率树辅助推理时，被试的强度估计模式不存在显著的因果方向差异；在使用隐藏嵌套集合关系的频率树辅助推理时，被试的强度估计模式在总体上存在显著的因果方向差异：更多被试在预防式问题上使用PPC估计因果强度，但在具体协变关系上不存在显著的因果方向差异。

以下先呈现两个实验，分别探讨反事实提问和能力提问条件下的图形促进效应，然后讨论频率树、因果方向、实验材料对强度估计的影响，并总结典型的因果强度估计模式。

2 实验 1：反事实提问条件下的图形促进效应研究

2.1 实验设计

2 × 2 × 3 完全随机实验设计，自变量是因果方向、频率树类型和协变关系，其中因果方向有两个水平：产生式、预防式；频率树类型有两个水平：提供嵌套集合关系频率树、隐藏嵌套集合关系频率树；协变关系有 3 个水平，水平 1：$\Delta P = 0.33$, PPC = 0.50；水平 2：$\Delta P = 0.33$, PPC = 0.83；水平 3：$\Delta P = 0.66$, PPC = 0.83，组合形成 12 个实验处理。因变量是被试使用各种作答模式的频数。

2.2 被试

2 × 2 × 3 完全随机实验设计共需 12 组被试，实验 1 的被试为 469 名大学生。为平衡被试专业背景对实验结果的影响，选择理科(化学、数学专业)、文科(文学、英语专业)和文理兼收(教育、财会、体育、音乐专业)的大学生，采取整群取样的方法，分班级利用晚自习和班会课时间在教室施测。每次施测有一半的被试使用隐藏嵌套集合关系的频率树，另一半被试使用提供嵌套集合关系的频率树辅助推理；在两部分被试中随机发放协变关系 1、2、3

的材料; 为避免被试之间的相互干扰, 每名被试的前后左右各有一个空位, 主试宣读指导语后要求被试独立观察协变关系材料并作答, 所有被试在15~30 分钟内完成了任务。

231 名大学生估计产生式协变关系的因果强度, 平均年龄 20.3 岁, 有 39(女 24)、41(女 23)、41(女 23)、37(女 21)、39(女 25)、34(女 20)人分别完成 ProCH1 (编码意义见表 3 的标注)、ProCO1、ProCH2、ProCO2、ProCH3、ProCO3 的强度估计; 238 名大学生估计预防式协变关系的因果强度, 平均年龄 19.4 岁, 有 38(女 18)、38(女 22)、39(女 22)、40(女 23)、41(女 26)、42(女 28)人分别完成 PreCH1、PreCO1、PreCH2、PreCO2、PreCH3、PreCO3 的因果强度估计。

2.3　材料和程序

每名被试拿到一张答题纸和一份协变关系材料。答题纸分两部分, 第一部分是指导语, 第二部分是被试作答区域。指导语采用 Liljeholm 和 Cheng (2009)的服药−头痛背景故事; 作答区域有 4 个环节:

(1) 要求被试观测协变关系材料。

协变关系材料共 30 页, 每页用文字和图形两种形式记载一名病人在服药前后的头痛情况(病人的头痛情况是图 1a、1b、1c、1d 之中的一种, 4 种情况互斥且完备, 以下将一名病人服药前后的头痛情况简称为“样例”)。为保证被试对目标原因的因果方向有明确的认识, 产生式协变关系中不出现药物消除头痛的样例(图 1b), 预防式协变关系中不出现药物导致头痛的样例(图 1a)。

样例数量、协变关系、估计模式的对应关系见表 2, 以产生式协变关系 2 为例, 实验材料包括 10 张图 1a、18 张图 1c、2 张图 1d, 对应 10 人服药前不头痛服药后头痛, 18 人服药前头痛服药后也头痛, 2 人服药前不头痛服药后也不头痛。协变关系部件中 a 表示服药后头痛的样例数, 等于 28 (10 张图 1a + 18 张图 1c); b 表示服药后不头痛的样例数, 等于 2 (0 张图 1b + 2 张图 1d); c 表示服药前头痛样例数, 等于 18 (0 张图 1b + 18 张图 1c); d 表示服药前不头痛样例数, 等于 12 (10 张图 1a + 2 张图 1d)。协变关系的 $\Delta P = a/(a + b) - c/(c + d) = 28/30 - 18/30 = 10/30 = 0.33$, 表示服药后比服药前增加的头痛人数比例; $PPC = \Delta P/(d/(c + d)) = (10/30)/(12/30) = 10/12 = 0.83$, 表示在服药前不头痛的病人中服药后头痛的比例; 相应地, 产生式协变关系 1 材料包含图 1a、1c、1d 各 10 张, $\Delta P = a/(a + b) - c/(c + d) = 20/30 - 10/30 = 10/30 = 0.33$, $PPC = \Delta P/(d/(c + d)) = (10/30)/(20/30) = 10/20 = 0.5$; 协变关系 3 的材料包含 20 张图 1a、6 张图 1c、4 张图 1d, $\Delta P = a/(a + b) - c/(c + d) = 26/30 - 6/30 = 20/30 = 0.67$, $PPC = \Delta P/(d/(c + d)) = (20/30)/(24/30) = 20/24 = 0.83$。

症状	是否服药	是否头痛
服用药物前不头痛		
服用药物后头痛		

a

症状	是否服药	是否头痛
服用药物前头痛		
服用药物后不头痛		

b

症状	是否服药	是否头痛
服用药物前头痛		
服用药物后头痛		

c

症状	是否服药	是否头痛
服用药物前不头痛		
服用药物后不头痛		

d

图 1　服药前后是否头疼样例

预防式协变关系 2 包括 10 张图 1b、2 张图 1c、18 张图 1d 分别对应 10 人服药前头痛服药后不头痛, 2 人服药前头痛服药后也头痛, 18 人服药前不头痛服药后也不头痛。协变关系部件中 a 表示服药后头痛的样例数, 等于 2 (0 张图 1a + 2 张图 1c); b 表示服药后不头痛的样例数, 等于 28 (10 张图 1b + 18 张图 1d); c 表示服药前头痛样例数, 等于 12 (10 张图 1b + 2 张图 1c); d 表示服药前不头痛样例数, 等于 18 (0 张图 1a+18 张图 1d)。协变关系的 $\Delta P = a/(a + b) - c/(c + d) = 2/30 - 12/30 = -10/30 = -0.33$ 表示服药后比服药前减少的头痛人数比例; $PPC = -\Delta P/(c/(c + d)) = -(-10/30)/(12/30) = 10/12 = 0.83$ 表示服药前头痛的病人中服药后不头痛的比例, 相应地, 预防式协变关系 1 材料包含图 1b、图 1c、图 1d 各 10 张, $\Delta P = a/(a + b) - c/(c + d) = 10/30 - 20/30 =$

表 2 实验刺激参数设置表

因果方向		样例数				协变关系部件				ΔP	PPC	P (E/C)	P (−E/C)
		图 1a	图 1b	图 1c	图 1d	a	b	c	d				
产生式	1	10	0	10	10	20	10	10	20	0.33	0.50	0.67	0.33
	2	10	0	18	2	28	2	18	12	0.33	0.83	0.93	0.07
	3	20	0	6	4	26	4	6	24	0.67	0.83	0.87	0.13
预防式	1	0	10	10	10	10	20	20	10	−0.33	0.50	0.33	0.67
	2	0	10	2	18	2	28	12	18	−0.33	0.83	0.07	0.93
	3	0	20	4	6	4	26	24	6	−0.67	0.83	0.13	0.87

注：P (E/C)指目标原因出现后效果出现的概率，是被试在产生式条件下的典型作答模式之一; P (−E/C)指目标原因出现后效果不出现的概率，是被试在预防式条件下的典型作答模式之一。

$-10/30 = -0.33$, $PPC = -\Delta P/(c/(c+d)) = -(-10/30)/(20/30) = 10/20 = 0.5$；预防式协变关系 3 材料包含 20 张图 1b、4 张图 1c、6 张图 1d, $\Delta P = a/(a + b) - c/(c + d) = 4/30 - 24/30 = -20/30 = -0.67$, $PPC = -\Delta P/(c/(c + d)) = -(-20/30)/(24/30) = 20/24 = 0.83$。

(2) 要求被试完善提供嵌套集合关系的频率树(图 2)或隐藏嵌套集合关系的频率树(图 3)。提供嵌套集合关系频率树共三层信息：第一层是协变关系中的样例总数；第二层是服药前头痛/不头痛样例数，其中服药前头痛样例数是预防式问题焦点集，服药前不头痛样例数是产生式问题焦点集；第三层是服药后头痛样/不头痛样例数。隐藏嵌套集合关系频率树共两层信息：第一层是协变关系中的样例总数；第二层是服药前不头痛服药后也不头痛、服药前不头痛服药后头痛、服药前头痛服药后也头痛、服药前头痛服药后不头痛的样例数。对同一个协变关系而言，提供嵌套集合关系频率树的第一层和第三层信息与隐藏嵌套集合关系频率树的第一层与第二层信息是相同的，提供嵌套集合关系频率树通过第二层信息强制被试注意焦点集，帮助被试明确数据间的嵌套集合关系。

(3) 要求被试做出因果强度估计。采用反事实提问，产生式协变关系的提问方式是：

现有 100 名病人在服用药物前不头痛，请问他们之中有______人在服用药物后会头痛？

预防式协变关系的提问方式是：

现有 100 名病人在服用药物前头痛，请问他们之中有______人在服用药物后不头痛？

(4) 要求被试在一个方框内用文字或算式描述得出强度估计的思维过程。

2.4 评判标准

依据被试的强度估计和在方框内描述的思维过程判定被试的作答模式，发现被试在产生式协变

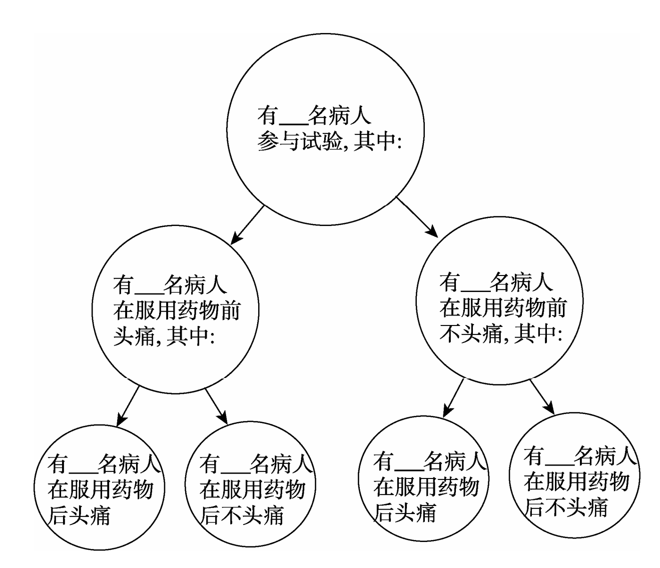


图 2 提供嵌套集合关系频率树

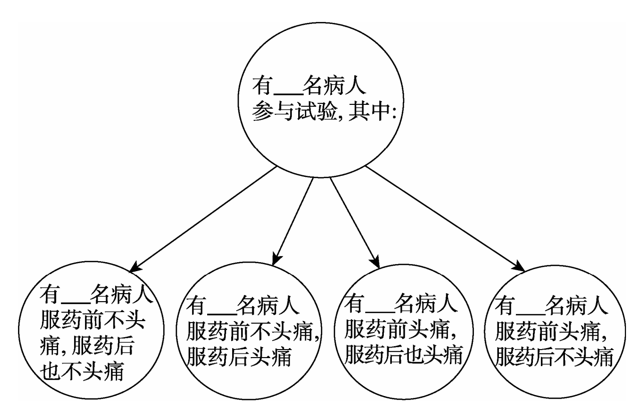


图 3 隐藏嵌套集合关系频率树

关系上较有代表性的作答模式有 ΔP、PPC 和 P(E/C)三种(P(E/C)指服药后头痛样例比例，在本研究中就是图 1a、图 1c 数量之和与样例总数 30 之商)，在预防式协变关系上较有代表性的作答模式有 ΔP、PPC 和 P (−E/C)三种(P (−E/C)指服药后不头痛样例比例，在本研究中就是图 1b、图 1d 数量之和与样例总数 30 之商)。以协变关系 1 为例，被判定为使用 ΔP 作答的强度估计为 33.3 或 30，且思维过程为

以下三者之一：(1)文字陈述：30 名病人中有 10 人服药前不头痛服药后头痛(预防式的文字陈述为：30 名病人中有 10 人服药前头痛服药后不头痛) (2)算式表达：(10/30) × 100, (3)算式表达：(20/30) × (1/2) × 100; 被判定为使用 PPC 作答的强度估计为 50, 且思维过程为以下二者之一：(1)文字陈述：20 名服药前不头痛的病人中有 10 人(或：一半)在服药后头痛(预防式的文字陈述为：20 名服药前头痛的病人中有 10 人(或：一半)在服药后不头痛); (2)算式表达：(10/20) × 100; 产生式协变关系上被判定为使用 P (E/C)作答、预防式协变关系上被判定为使用 P (−E/C)作答的强度估计为 67 或 70, 且思维过程为以下二者之一：(1)文字陈述：大多数病人在服药后头痛(预防式的文字陈述为：大多数病人在服药后不头痛); (2)算式陈述(20/30) × 100。

2.5　实验结果

被试在各协变关系上的作答模式见表 3。

检验研究假设 1。由于 PPC 值是强度估计的极大似然估计，所以将合理的因果强度估计界定为使用 PPC 值估计协变关系的因果强度。将被试在协变关系相同但频率树类型不同的 6 组处理(产生式的 ProCH1 与 ProCO1、ProCH2 与 ProCO2、ProCH3 与 ProCO3; 预防式的 PreCH1 与 PreCO1、PreCH2 与 PreCO2、PreCH3 与 PreCO3)上的作答模式区分为使用 PPC 作答和使用非 PPC 作答(以 ProCH1 与 ProCO1 组为例，被试在 ProCH1 问题上使用 PPC 作答的频数为 6, 使用非 PPC 作答的频数为 26 + 5 + 2 = 33; 在 ProCO1 问题上使用 PPC 作答的频数为 33, 使用非 PPC 作答的频数为 5 + 2 + 1 = 8), 对 6 组数据分别做作答模式(使用 PPC 作答、使用非 PPC 作答)是否受频率树类型(提供嵌套集合关系、隐藏嵌套集合关系)影响的独立性卡方检验，所有检验结果都显示：与隐藏嵌套集合关系频率树条件相比，有显著更多的被试在提供嵌套集合关系频率树条件下使用 PPC 估计因果强度，双尾检验卡方值分别为 ProCH1 与 ProCO1 组 $\chi^2_{(1)} = 33.03$, $p < 0.001$, $\Phi = 0.64$; ProCH2 与 ProCO2 组 $\chi^2_{(1)} = 28.61$, $p < 0.001$, $\Phi = 0.61$; ProCH3 与 ProCO3 组 $\chi^2_{(1)} = 33.79$, $p < 0.001$, $\Phi = 0.68$; PreCH1 与 PreCO1 组 $\chi^2_{(1)} = 21.11$, $p < 0.001$, $\Phi = 0.53$; PreCH2 与 PreCO2 组 $\chi^2_{(1)} = 11.11$, $p = 0.004$, $\Phi = 0.38$; PreCH3 与 PreCO3 组 $\chi^2_{(1)} = 28.93$, $p < 0.001$, $\Phi = 0.59$; 实验结果证实研究假设 1。

检验研究假设 2。在产生式条件下，将使用隐藏嵌套集合关系频率树的 3 个处理(ProCH1、ProCH2、ProCH3)和提供嵌套集合关系频率树的 3 个处理(ProCO1、ProCO2、ProCO3)上的作答模式区分为使用 PPC 作答和使用非 PPC 作答，分别做作答模式是否受协变关系(协变关系 1、2、3)影响的独立性卡方检验，结果显示在使用提供嵌套集合关系的频率树辅助推理时，被试的作答模式在协变关系间不存在显著差异；在使用隐藏嵌套集合关系的频率树辅助推理时，被试的作答模式在协变关系

表 3　被试作答模式描述统计(反事实提问)

协变关系	产生式作答模式						预防式作答模式					
	编码	指标	ΔP	PPC	P (E/C)	其他	编码	指标	ΔP	PPC	P (−E/C)	其他
1	ProCH1	频数	26	6	5	2	PreCH1	频数	26	8	3	1
		百分比	67%	15%	13%	5%		百分比	68%	21%	8%	3%
	ProCO1	频数	5	33	2	1	PreCO1	频数	6	28	2	2
		百分比	12%	80%	5%	2%		百分比	16%	74%	5%	5%
2	ProCH2	频数	25	12	2	2	PreCH2	频数	19	17	3	0
		百分比	61%	29%	5%	5%		百分比	49%	44%	8%	0%
	ProCO2	频数	4	33	0	0	PreCO2	频数	7	32	1	0
		百分比	11%	89%	0%	0%		百分比	18%	80%	3%	%
3	ProCH3	频数	33	3	2	1	PreCH3	频数	29	9	1	2
		百分比	85%	8%	5%	3%		百分比	71%	22%	2%	5%
	ProCO3	频数	6	24	2	2	PreCO3	频数	1	34	3	4
		百分比	18%	71%	6%	6%		百分比	2%	81%	7%	10%

注：为方便陈述实验结果，编码表示实验 1 的 12 个处理。编码含 4 个部分，其中用 Pro 指代因果方向中的产生式(Product)，用 Pre 指代因果方向中的预防式(Prevent); C 指代提问方式中的反事实提问(Counter-factual), A 指代提问方式中的能力提问(Ability, 见表 4); H 指代频率树类型中的隐藏嵌套集合关系频率树(Hide), O 指代频率树类型中的提供嵌套集合关系频率树(Offer); 数值指代协变关系 1、2、3; 实验 2 的 12 个处理采用相同的编码系统。

间差异显著，$\chi^2_{(1)} = 6.515$, $p = 0.038$, $\Phi = 0.23$，两两比较结果显示作答模式差异出现在协变关系 2 与协变关系 3 之间，有更多被试在协变关系 2 上使用 PPC 估计因果强度，矫正 $\chi^2_{(1)} = 4.7730$, $p = 0.029$, $\Phi = 0.24$；对被试在预防式条件下的作答模式做相同的检验，结果显示在使用提供嵌套集合关系频率树辅助推理时(对应实验处理 PreCO1、PreCO2、PreCO3)，被试的作答模式在协变关系间不存在显著差异；在使用隐藏嵌套集合关系频率树辅助推理时(对应实验处理 PreCH1、PreCH2、PreCH3)，被试的作答模式在协变关系间差异显著：$\chi^2_{(1)} = 6.21$, $p = 0.045$, $\Phi = 0.23$，两两比较结果显示作答模式差异出现在协变关系 2 与协变关系 1 之间，矫正 $\chi^2_{(1)} = 3.49$, $p = 0.035$, $\Phi = 0.21$，有更多被试在协变关系 2 上使用 PPC 估计因果强度；实验结果证实研究假设 2。

检验研究假设 3。对 ProCH1 与 PreCH1、ProCO1 与 PreCO1、ProCH2 与 PreCH2、ProCO2 与 PreCO2、ProCH3 与 PreCH3、ProCO3 与 PreCO3 共 6 组数据分别做作答模式(使用 PPC 作答、使用非 PPC 作答)是否受因果方向(产生式、预防式)影响的独立性卡方检验，6 组数据的卡方检验结果均未达到显著，显示就单个协变关系而言，被试的作答模式不受因果方向的影响。综合提供嵌套集合关系产生式 ProCO1、ProCO2、ProCO3 和预防式 PreCO1、PreCO2、PreCO3 的作答模式，做作答模式是否受因果方向影响的独立性卡方检验，结果显示差异并不显著；综合隐藏嵌套集合关系产生式 ProCH1、ProCH2、ProCH3 和预防式 PreCH1、PreCH2、PreCH3 的作答模式，做作答模式是否受因果方向影响的独立性卡方检验，结果显示差异显著：$\chi_{(1)}{}^2 = 4.416$, $p = 0.042$, $\Phi = 0.132$，有更多的被试在预防式条件下使用 PPC 估计因果强度。检验结果显示被试作答模式的因果方向差异在总体上是存在的，但就单个协变关系而言，作答模式不存在显著的因果方向差异。结果证实研究假设 3。

实验 1 采用强制被试关注焦点集信息的方式为被试提供嵌套集合关系，发现了效果明显的图形促进效应。但这个实验使用的反事实提问也有引导被试注意嵌套集合关系的作用：如产生式反事实提问中“现有 100 名病人在服用药物前不头痛……服药后头痛”的措辞可能引导被试注意服药前不头痛的病人中服药后头痛的比例，进而获得数据间的嵌套集合关系。因此被试对嵌套集合关系的理解可能是由提供嵌套集合关系的频率树导致的，也可能是由反事实提问导致的。为消除提问方式对被试的引导作用，实验 2 使用能力提问，探讨在提问方式上不引导被试关注嵌套集合关系时，频率树是否影响被试因果推理作答模式。

3 实验 2：能力提问条件下的图形促进效应研究

3.1 实验设计

同实验 1。

3.2 被试

大学生被试 463 人，采用与实验 1 相同的方法平衡无关因素的影响，被试来自理科(数学)、文科(英语)和文理兼收(教育、财会)专业。其中 238 人估计产生式协变关系的因果强度，平均年龄 18.4 岁，有 43(女 24)、39(女 24)、41(女 26)、38(女 23)、39(女 21)、38(女 21)人分别完成 ProAH1、ProAO1、ProAH2、ProAO2、ProAH3、ProAO3 的强度估计；225 人估计预防式协变关系的因果强度，平均年龄 18.6 岁，有 36(女 18)、38(女 23)、38(女 21)、38(女 22)、37(女 15)、38(女 25)人分别完成 PreAH1、PreAO1、PreAH2、PreAO2、PreAH3、PreAO3 的强度估计。

3.3 材料和程序

实验材料同实验 1，只是答题纸的提问部分改用王墨耘和傅小兰(2004b)使用的能力提问，产生式能力提问方式是：

该药物导致病人头痛症状的能力是：________。(使用 0~100 的数字描述该药物导致病人头痛症状的能力：0 表示该药物完全不会导致病人头痛，100 表示该药物一定会导致病人头痛，你认为该药物导致病人头痛的能力越强就给出越大的数值)

预防式能力提问方式与产生式相同，只是将所有的“导致”修改为“消除”。

实验程序同实验 1，所有被试在 12~30 分钟内完成了任务。

3.4 评判标准

同实验 1。

3.5 实验结果

被试在各协变关系上的作答模式见表 4。

检验研究假设 1。与实验 1 相同，将被试在协变关系相同但频率树不同的 6 组处理(产生式的 ProAH1 与 ProAO1、ProAH2 与 ProAO2、ProAH3 与 ProAO3；预防式的 PreAH1 与 PreAO1、PreAH2 与 PreAO2、PreAH3 与 PreAO3)上的作答模式整理

表 4　被试作答模式描述统计(能力提问)

协变关系	产生式作答模式						预防式作答模式					
	编码	指标	ΔP	PPC	P (E/C)	其他	编码	指标	ΔP	PPC	P (−E/C)	其他
1	ProAH1	频数	19	6	11	7	PreAH1	频数	11	10	8	7
		百分比	44%	14%	26%	16%		百分比	31%	28%	22%	19%
	ProAO1	频数	6	19	7	7	PreAO1	频数	1	31	2	4
		百分比	15%	49%	18%	18%		百分比	3%	82%	5%	10%
2	ProAH2	频数	11	10	8	12	PreAH2	频数	9	14	8	7
		百分比	27%	24%	20%	29%		百分比	24%	37%	21%	18%
	ProAO2	频数	2	23	3	10	PreAO2	频数	1	27	6	4
		百分比	5%	61%	8%	26%		百分比	3%	71%	16%	10%
3	ProAH3	频数	25	0	8	6	PreAH3	频数	20	5	6	6
		百分比	64%	0%	21%	15%		百分比	54%	14%	16%	16%
	ProAO3	频数	0	20	8	10	PreAO3	频数	1	22	9	6
		百分比	0%	53%	21%	26%		百分比	3%	58%	24%	16%

为使用 PPC 作答和使用非 PPC 作答两类，对 6 组数据分别做作答模式是否受频率树类型影响的独立性卡方检验，所有检验结果都显示：与隐藏嵌套集合关系条件相比，有显著更多的被试在提供嵌套集合关系条件下使用PPC估计因果强度，双尾卡方检验结果分别为:ProAH1 与 ProAO1 组 $\chi^2_{(1)} = 11.66$, $p = 0.001$, $\Phi = 0.38$; ProAH2 与 ProAO2 组 $\chi^2_{(1)} = 10.59$, $p = 0.001$, $\Phi = 0.37$; ProAH3 与 ProAO3 组 $\chi^2_{(1)} = 27.73$, $p < 0.001$, $\Phi = 0.60$; PreAH1 与 PreAO1 组 $\chi^2_{(1)} = 21.66$, $p < 0.001$, $\Phi = 0.54$; PreAH2 与 PreAO2 组 $\chi^2_{(1)} = 8.95$, $p = 0.003$, $\Phi = 0.34$; PreAH3 与 PreAO3 组 $\chi^2_{(1)} = 16.03$, $p < 0.001$, $\Phi = 0.46$; 实验结果证实研究假设 1。

检验研究假设 2。在产生式条件下，将被试在隐藏嵌套集合关系的 3 个处理(ProAH1、ProAH2、ProAH3)和提供嵌套集合关系 3 个处理(ProAO1、ProAO2、ProAO3)上的作答模式区分为使用 PPC 作答和使用非PPC作答，分别做作答模式是否受协变关系影响的独立性卡方检验，结果显示在使用提供嵌套集合关系的频率树辅助推理时，被试的作答模式在协变关系间不存在显著差异；在使用隐藏嵌套集合关系的频率树辅助推理时，被试的作答模式在协变关系间差异显著，$\chi^2_{(2)} = 10.56$, $p = 0.005$, $\Phi = 0.29$，两两比较结果显示作答模式差异出现在协变关系 2 与协变关系 3 之间，矫正 $\chi^2_{(1)} = 8.76$, $p = 0.003$, $\Phi = 0.33$，有更多被试在协变关系 2 上使用 PPC 作答，作答模式差异在协变关系 1 与协变关系 3 之间的差异边缘显著：矫正 $\chi^2_{(1)} = 3.99$, $p = 0.046$, $\Phi = 0.22$，有更多被试在协变关系 1 上使用 PPC 作答，这可能是因为协变关系 3 上使用 PPC 估计因果强度的频数为零所致；对被试在预防式条件下的作答模式做同样的独立性卡方检验，结果显示在使用提供嵌套集合关系的频率树辅助推理时(对应实验处理 PreAO1、PreAO2、PreAO3)，被试的作答模式在协变关系间不存在显著差异；在使用隐藏嵌套集合关系的频率树辅助推理时(对应实验处理 PreAH1、PreAH2、PreAH3)，被试的作答模式在协变关系间差异边缘显著，$\chi^2_{(2)} = 5.36$, $p = 0.069$, $\Phi = 0.22$，两两比较结果显示作答模式差异出现在协变关系 2 与协变关系 3 之间，矫正 $\chi^2_{(1)} = 4.23$, $p = 0.04$, $\Phi = 0.24$，有更多被试在协变关系 2 上使用 PPC 作答；实验结果证实研究假设 2。

检验研究假设 3。对 ProAH1 与 PreAH1、ProAO1 与 PreAO1、ProAH2 与 PreAH2、ProAO2 与 PreAO2、ProAH3 与 PreAH3、ProAO3 与 PreAO3 共 6 组数据分别做作答模式是否受因果方向影响的独立性卡方检验，结果显示 ProAO1 与 PreAO1 组 $\chi^2_{(1)} = 9.128$, $p = 0.003$, $\Phi = 0.34$，有更多被试在预防式问题上使用 PPC 作答；其余卡方检验结果均未达到显著，显示就大多数协变关系而言，被试的作答模式不受因果方向的影响。综合提供嵌套集合关系产生式 ProAO1、ProAO2、ProAO3 和预防式 PreAO1、PreAO2、PreAO3 的作答模式，做作答模式是否受因果方向影响的独立性卡方检验，结果显示差异显著：$\chi^2_{(1)} = 6.427$, $p = 0.011$, $\Phi = 0.168$，暗示总体上看，有显著更多的被试在预防式条件下使用 PPC 估计因果强度；综合隐藏嵌套集合关系产生式 ProAH1、ProAH2、ProAH3 和预防式 PreAH1、

PreAH2、PreAH3 的作答模式，做作答模式是否受因果方向影响的独立性卡方检验，结果显示差异显著：$\chi^2_{(1)} = 4.464$, $p = 0.011$, $\Phi = 0.166$，暗示总体上看，有显著更多的被试在预防式条件下使用 PPC 估计因果强度。绝大部分实验结果都证实研究假设 3，只在提供嵌套集合关系的协变关系 1 上有更多被试在预防式条件下使用PPC估计因果强度，与研究假设不符，这可能也是在提供嵌套集合关系条件总体上存在因果方向差异的原因。由于这个差异只出现在单个实验的单个协变关系上，对这个结果只报告不讨论。

综合两个实验的结果，提供嵌套集合关系的频率树在反事实提问和能力提问条件下都能提高被试推理的合理性；协变关系差异比较微弱但稳定出现在隐藏嵌套集合关系的实验处理中，因果方向差异稳定出现在隐藏嵌套集合关系的实验处理中。

4 讨论

本研究使用提供/隐藏嵌套集合关系的频率树探究频率树对因果强度估计的促进作用，结果显示不管是在反事实还是能力提问条件下，提供嵌套集合关系的频率树都能提高被试使用 PPC 作答的概率，只有少数使用隐藏嵌套集合关系频率树的被试使用PPC作答，说明提供嵌套集合关系的频率树才具有促进作用，实验结果与贝叶斯推理和三段论推理领域的研究结果是一致的。本研究的新发现是促进推理合理性的图形应随推理任务的不同而发生变化，使用焦点集提供嵌套集合关系的频率树是提升因果推理合理性的合适图形。Vallée-Tourangeau 等(2008)将贝叶斯推理研究中的频率树应用于因果推理研究，实验结果并不能证明大多数被试使用 PPC 估计因果强度。本研究使用的提供嵌套集合关系频率树的构建方式是总样例——焦点集(服药前头痛/不头痛样例数)——服药后头痛/不头痛样例数，发现大部分借助该频率树辅助推理的被试使用 PPC 估计因果强度(人数比例在 48.72%～89.19%之间，平均 70.6%)。同时，不管是反事实还是在能力提问条件，不管是产生式还是在预防式协变关系条件，也不管协变关系的 ΔP 值和 PPC 值，与借助隐藏嵌套集合关系的频率树辅助推理相比，有显著更多的被试在借助提供嵌套集合关系频率树辅助推理时使用PPC估计因果强度，暗示提供嵌套集合关系的频率树可以有效提高被试的因果推理合理性，干预效果具有普遍意义。这一促进效应的原因可能是因为提供嵌套集合关系频率树形象化地表征了焦点集样例与目标原因导致效果样例之间的关系，减轻了被试使用PPC估计因果强度的认知负荷，这也被认为是贝叶斯推理领域图形促进效应的原因(Hoffrage et al., 2015)，后续研究可以将提供嵌套集合关系频率树应用于提升青少年因果推理能力的教育实践和纠正成人的因果推理偏差，拓展图形促进效应研究成果的外部效度。

Buehner 等(2003)发现反事实提问这一干预手段能有效提高被试使用 PPC 作答的概率，但 Perales 和 Shanks (2008)的实验 3b 显示被试在完成反事实提问任务时并不使用PPC估计因果强度，本研究的结果也不支持反事实提问一定能提高被试使用 PPC 估计因果强度的概率：比较实验 1 和实验 2 的结果，与能力提问相比，反事实提问在 3 个提供嵌套集合关系的实验条件上促使更多被试使用 PPC 估计因果强度：ProCO1 和 ProAO1 组 $\chi^2_{(1)} = 8.87$, $p = 0.003$, $\Phi = 0.33$; ProCO2 和 ProAO2 组 $\chi^2_{(1)} = 8.14$, $p = 0.004$, $\Phi = 0.33$; PreCO3 和 PreAO3 组 $\chi^2_{(1)} = 5.05$, $p = 0.025$, $\Phi = 0.25$；在其他提供嵌套集合关系实验条件和所有隐藏嵌套集合关系实验条件上，反事实提问并未表现出促进效应。综合 3 个协变关系的数据比较频率树类型和提问方式的影响力，发现在产生式(图 4a)和预防式(图 4b)条件上，频率树类型的变化都导致被试使用 PPC 估计因果强度的频数发生显著变化；提问方式的变化也导致被试使用 PPC 估计因果强度的频数发生变化，反事实提问比能力提问导致更多被试使用PPC估计因果强度，但只在提供嵌套集合关系的产生式条件导致显著差异。综合起来，本研究认为频率树类型和提问方式共同影响被试的因果强度估计模式，被试在提供嵌套集合关系频率树和反事实提问的条件组合上使用 PPC 估计因果强度的比例最高，但就促进效果的普遍性而言，提问方式不及频率树类型。后续研究可以进一步探讨干预被试思维过程的具体方法，继续为提高被试的因果推理合理性提供可操作的建议。

协变关系的呈现方式有序列呈现(Perales & Shanks, 2007; Luhmann & Ahn, 2011; Lu, Rojas, Beckers, & Yuille, 2016; Bramley, Dayan, Griffiths, & Lagnado, 2017)和集中呈现(Buehner et al., 2003; Lu et al., 2008; Yeung & Gtriffiths, 2015)两种，序列呈现指逐个呈现目标原因是否导致结果的样例，一个样例呈现后被后续样例覆盖，被试不能记录样例信息；集中呈现指使用文字、表格和图形等形式概

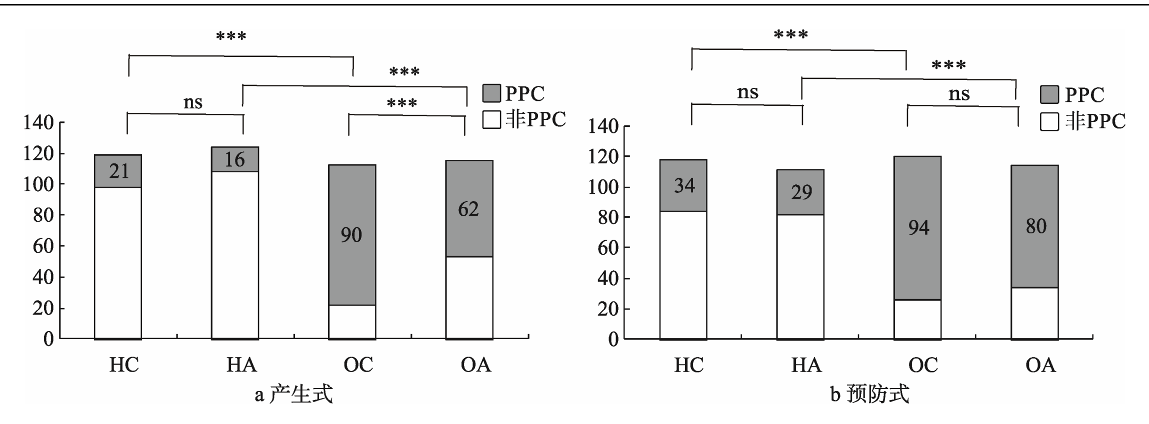


图 4　分实验条件的估计模式条形图

注: X 轴编码 H 指隐藏嵌套集合关系, O 指提供嵌套集合关系, A 指能力提问, C 指反事实提问; ***表示显著性水平 $p < 0.001$, ns 表示无显著性差异。

括呈现全部因果样例。被试在序列呈现范式上的强度估计常因目标原因导致效果的样例在序列中的位置变化而发生改变, 表现为近因效应或首因效应(Luhmann & Ahn, 2011; Lu et al., 2016); 集中呈现方式则存在一些威胁因果推理根本假设的问题, 如不能保证背景原因在不同组(服药组和不服药组)导致相同数量的效果, 不能保证被试对目标原因的作用方向(预防式或产生式)有统一的认识等, 这些问题可能导致被试认为协变关系与因果推理没有关系(Lu et al., 2008)。本研究使用样例册提供协变关系信息, 通过服药前、后的设计保证背景原因效力的一致性; 通过产生式协变关系不出现图 1b, 预防式协变关系不出现图 1a 保证被试对目标原因的作用方向只能形成一种认识; 通过样例册和频率树保证被试的作答不受记忆负荷的影响。从本质上看样例册是一种集中呈现信息的方式, 它相较于其他集中呈现方式的优势包括: (1)综合了序列呈现和集中呈现的优点, 保证被试对背景原因的效力和目标原因的因果方向有正确认识; (2)每页呈现一个样例的形式模拟了人类被试在日常生活中接触因果事件的真实过程, 是生态化的协变关系呈现方式。事实上, 绝大多数以动物(Dwyer & Waldmann, 2016)和人类儿童(Vul, Goodman, Griffiths, & Tenenbaum, 2014; McCormack, Bramley, Frosch, Patrick, & Lagnado, 2016)为被试的因果推理研究都采用逐个呈现样例的方式(近年综述见 Blaisdell, Sawa, Leising, & Waldmann, 2006 和 Gopnik, 2012); (3)样例之间互斥且完备, 使频率树的构建成为可能。鉴于样例册有助于消除被试对因果关系的模糊认识, 后续研究可将样例册应用于检验启发式因素对因果推理的影响力, 这些因素包括总体取样和实质取样(Liljeholm & Cheng, 2009)、协变关系中 a 的大小和目标原因导致效果样例的出现顺序(Luhmann & Ahn, 2011)等。

与大多数频率促进领域的研究为被试提供完整频率树不同, 本研究提供不完整频率树, 要求被试在整理样例册信息后完善频率树。与直接获得完整图形表征相比, 完善不完整图形需要更多的认知加工, 但在大多数条件下无损于被试的推理表现(Cosmides & Tooby, 1996, 实验 4 的维尼图条件; 史滋福等 2006, 实验 2; Brase, 2009, 实验 2), 在频率格条件下还可能提升被试的推理表现(Cosmides & Tooby, 1996, 实验 4), Cosmides 和 Tooby (1996)认为完善图形的促进作用可能是因为完善图形需要被试提供问题涉及的信息: 如一个使用最为广泛的贝叶斯推理问题是“检测结果为阳性的病人中有多少人罹患癌症”, 而完善图形的任务包括填充检验结果为阳性的人数和罹患癌症的人数, 被试因此只需进行简单的除法运算就能得到正确结论, 并不需要进行复杂的贝叶斯推理(P35)。可能正是基于这样的原因, 不完整频率树在推理能力培养的过程中能发挥重要作用(Sedlmeier & Gigerenzer, 2001)。但是以上证据都来自于贝叶斯推理或概率推理领域而非因果推理领域, 完整/不完整频率树对被试因果推理的影响是否存在差异尚未得到直接检验。结合以往研究和本研究的实验结果, 我们形成以下有待检验的假设: (1)与直接获得完整频率树相比, 被试基于样例册信息完善频率树需要更多的认知加工, 但由于绝大多数成人被试都能正确完善频率树, 获得与完整频率树条件相同的图形表征, 因此在多数条件下不完整频率树并不影响被试的因果推理表现; (2)在借助提供嵌套集合关系频率树辅助推理时, 提问方式(反事实提问、能力提问)与频率树性

质(完整、不完整)交互影响被试的因果推理：在完整频率树条件，反事实提问和能力提问导致的因果推理差异较小；在不完整频率树条件，由于被试完善频率树的部分信息与反事实提问采用的措辞(“现有 100 名病人在服用药物前不头痛……服药后头痛”)相吻合，反事实提问可能比能力提问导致更多的被试使用 PPC 估计因果强度。后续研究可进一步验证这些假设。

与王墨耘和傅晓兰(2004b)、Perales 和 Shanks (2008)的研究结果一样，本研究发现了 3 种典型的作答模式：ΔP、PPC、P (E/C) (产生式)或 P (−E/C) (预防式)，作答模式随实验条件的变化而变化，如在反事实提问 + 提供嵌套集合关系频率树条件，大部分被试使用 PPC 作答；但在能力提问+隐藏嵌套集合关系频率树条件并未发现占优势的作答模式。鉴于作答模式的多样性，本研究认为使用一种作答模式来预测所有被试在所有条件下的作答表现是不合适的，接纳多种作答模式应该成为后续模型建构的方法之一。包含多个作答模式的模型既可以解释被试的因果强度估计在 PPC 不变时随 ΔP 变化而变化，在 ΔP 不变时随 PPC 变化而变化的现象，也可解释实验任务差异、实验条件差异造成的作答模式差异。Perales 和 Shanks (2007)使用加权后的 ΔP 模型和 PPC 模型构建的 HPN 模型是一个有益的尝试，但这类研究还很少，需要构建更多包含多个作答模式的模型来解释和预测人类被试的因果推理表现。

5 研究结论

(1) 明确数据之间的嵌套集合关系能极大地提高被试使用PPC估计因果强度的概率，提供焦点集的频率树可以帮助被试明确嵌套集合关系；

(2) 频率树类型和提问方式共同影响被试的因果强度估计模式，提供嵌套集合关系频率树+反事实提问的条件组合促使最多被试使用 PPC 估计因果强度，但提问方式促进效果的普遍性不及频率树类型。

(3) 在借助隐藏嵌套集合关系的频率树辅助推理时，观察到一个微弱但稳定的协变关系差异：更多被试在 ΔP 小且 PPC 大的协变关系上使用 PPC 估计因果强度，同时观察到一个稳定的因果方向差异：在不区分协变关系时，更多被试在预防式问题上使用PPC估计因果强度，但该差异在单个协变关系上并不显著。

致谢：感谢江西科技师范大学材料与机电学院刘庭芝博士、经管学院刘伟老师、数信学院易云辉老师，研究生魏臻、龙兵、陈素清在收集整理数据和论文撰写过程中提供的帮助。

参 考 文 献

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology, 23*, 369–381.

Brase, G. L. (2014). The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology, 26*, 81–97.

Buehner, M. J., Cheng, P. W., & Clliford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 1119–1140.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing neurath’s ship: Approximate algorithms for online causal learning. *Psychological Review, 124*, 301−338.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology, 76*, 30–77.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*, 1020–1022.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Cheng, P. W., & Buehner, M. J. (2012). Causal learning. In K. J. Holyoak & R. G. Morrison (Eds), *The oxford handbook of thinking and reasoning* (pp.210–233). New York: Oxford University Press.

Cheng, P., Liljeholm, M., & Sandhofer, C. (2013). Logical consistency and objectivity in causal learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 2034–2039). Austin, TX: Cognitive Science Society.

Carroll, C., Cheng, P. W., & Lu, H. (2013). Inferential dependencies in causal inference: A comparison of belief-distribution and associative approaches. *Journal of Experimental Psychology: General, 142*, 845–863.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58*, 1–73.

Dwyer, D. M., & Waldmann, M. R. (2016). Beyond the information (not) given: Representations of stimulus absence in rats (*Rattus norvegicus*). *Journal of Comparative Psychology, 130*(3), 192–204.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*, 1623–1627.

Gigerenzer G., & Hoffrage U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychology Review, 102*, 684−704.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.

Hill, W. T., & Brase, G. L. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *The Quarterly Journal of Experimental Psychology,*

65, 2343–2368.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology, 62*, 135–163.

Hoffrage, U., Hafenbrädl, S., & Bouquet, C. (2015). Natural frequencies facilitate diagnostic inferences of managers. *Frontiers in Psychology, 6*, 642.

Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs:General and Applied, 7*, 1–17.

Luhmann, C. C., & Ahn, W. K. (2011). Expectations and interpretations during causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 568–587.

Liljeholm, M., & Cheng, P. W. (2009). The influence of virtual sample size on confidence and causal-strength judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 157–172.

Liu, Y. L., Chen, S. P., & Hu, Z. J. (2015). A research on virtual sample alter causal strength estimate. *Journal of Psychological Science, 38*(1), 146–151.

[刘雁伶, 陈水平, 胡竹菁. (2015). 实质取样对因果强度估计的影响. *心理科学, 38*(1), 146–151.]

Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. M. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 260–276). New York: Oxford University Press.

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? Critique of Cheng (1997). *Psychological Review, 107*, 195–212.

Lu, H., Yuille, A., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955–984.

Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science, 40*, 404–439.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology, 141*, 1−22.

Oaksford, M., & Chater, N. (2010).*Cognition and conditionals: Probability and logic in human thinking.* New York: Oxford University Press.

Powell, D., Merrick, M. A., Lu, H. J., & Holyoak, K. J. (2016). Causal competition based on generic priors. *Cognitive Psychology, 86*, 62–86.

Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletiin & Review, 14*, 577–596.

Perales, J. C., & Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1482–1494.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinformecement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64−99). New York: Appleton-Century-Crofts.

Sedlmeier, P. & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*, 380–400.

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2015). How to train your Bayesian: A problem-representation transfer rather than a format-representation shift explains training effects. *The Quarterly Journal of Experimental Psychology, 68*, 1–9.

Shi, Z. F., Qiu, J, & Zhang, Q. L. (2006). Facilitating effect of transparent nested-sets relations on Bayesian reasoning. *Acta Psychologica Sinica, 38*, 833−840.

[史滋福, 邱江, 张庆林. (2006). 明确嵌套集合关系对贝叶斯推理的促进效应. *心理学报, 38*, 833–840.]

Thompson, V., & Evans, J. St. B. T. (2012). Belief bias in informal reasoning. *Thinking and Reasoning, 18*, 278–310.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38*, 599–637.

Vallée-Tourangeau, G., Abadie, M., & Vallée-Tourangeau, F. (2015). Interactivity fosters Bayesian reasoning without instruction. *Journal of Experimental Psychology: General, 144*, 581–603.

Vallée-Tourangeau, F., Payton, T., & Murphy, R. A. (2008). The impact of presentation format on causal inferences. *European Journal of Cognitive Psychology, 20*, 177–194.

Wang, M. Y., & Fu, X. L. (2004a). Causal judgments in the concentrative presentation of information with different external representations. *Acta Psychologica Sinica, 36*, 298–306.

[王墨耘, 傅小兰. (2004a). 用不同外部表征方式集中呈现信息条件下的因果力判断. *心理学报, 36*, 298–306.]

Wang, M. Y., & Fu, X. L. (2004b). A test of the power PC theory with the paradigm of comparing causal power. *Acta Psychologica Sinica, 36*, 160–167.

[王墨耘, 傅小兰. (2004b). 因果力比较范式下对效力 PC 理论的检验. *心理学报, 36*, 160–167.]

Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology, 76*, 1–29.

The promotion of frequency tree type and questioning format on causal strength estimation

LIU Yanling[1]; CHEN Jun[2]; SHEN Youtian[3]; HU Zhujing[3]

([1] *Department of Education, Jiangxi Science and Technology Normal University, Nanchang* 330038, *China*)

([2] *Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, *China*)

([3] *Department of Psychology, Jiangxi Normal University;*

Key Laboratory of Psychology and Cognition Science, Ministry by Jiangxi Province, Nanchang 330022, *China*)

Abstract

There are lots of evidences showing that participant's performance on Bayesian inference, syllogistic reasoning and probability reasoning could be promoted by cumulative frequency tree. However, very few study focuses on the promotion effect of frequency tree on causal reasoning. This study carried out two experiments to investigate the effect of frequency tree on causal strength inference. The research hypotheses include: (a) Frequency tree featuring a explicit nest-sets structure (ENS) can improve the rationality of participant's reasoning, while the frequency tree featuring a concealed nest-sets structure (CNS) can't improve rationality of reasoning; (b) Participants estimate the causal strength of different contingencies by different modes in experimental treatment which used frequency tree featuring a CNS; and (c) There are more participants estimate the causal strength by Power–PC model in preventive contingency rather than in productive contingency.

2 (Frequency tree, level 1: featuring a ENS, level 2: featuring a CNS) × 2 (causal direction, level 1: productive, level 2: preventive) × 3 (contingency, level 1: ΔP = 0.33 and Power – PC = 0.5; level 2: ΔP = 0.33 and Power – PC = 0.83; level 3: ΔP = 0.67 and Power – PC = 0.83) completely random design were used in two experiments. 469 undergraduate students participated in Experiment 1 which adopted counter–factual question, and 463 undergraduate students participated in Experiment 2 which adopted ability question. Contingency was offered by a booklet which contains 30 pages, and each page presents one sample related to the causality. Participant completed a frequency tree based on contingency, and estimated the causal strength of contingency individual. The frequency tree featuring a ENS consists of three types of information: the number of total samples, the number of samples in focus set, and the number of samples that represent effect emerge or not, while frequency tree featuring a CNS consists of the number of total samples and samples that represent effect emerge or not.

The study found that (a) There are three common models of causal reasoning: Δ*p*, Power–PC and P (E/C) for productive contingency (or P(-E/C) for preventive contingency), the most popular model changes with different experiment treatments; (b) 70.06 % of participants estimate causal strength by Power–PC model when they used frequency tree featuring a ENS, and only a few participants (about 21.28 %) estimate causal strength by Power–PC model when they used frequency tree featuring a CNS; (c) The type of frequency tree and the format of question have combining influence on causal strength evaluation, and the type of frequency tree have more influences on strength evaluation than the format of question; (d) Both contingency effect and causal direction effect are present from the experimental treatment which used frequency tree featuring a CNS. Experiment results significantly support research hypotheses (a), (b) and (c).

These results indicate that frequency facilitating effect depends on supply nest-sets structure or not, whether in counter–factual question treatment or in ability question treatment. According to above two experiments, it is suggested that participant tends to make rational inference when they use frequency tree featuring a ENS or they were questioned by counter–factual format.

Key words causal inference; promotion effect; frequency tree; question format